COMPUTER SCIENCE CS3753
Assignment #7

Points: 60   **Weight**: 2%

Due: Friday, November 23, 2018 at 11:55 pm in BlackBoard

Note: Late assignment will not be accepted without instructor's pre–approval.

Hand in a Jupyter notebook**, yourNmae-homework07.ipynb**, with appropriate markdown cells for descriptions and comments. Write Python code to solve each of the following questions. ***This homework must be completed individually.***

1. [20] Assume that the annual rate of return for a stock is a normal random variable with mean μ and standard deviation σ.

    (a) Write a python function buyshares that takes μ, σ, current price p, expected profit b in dollars after one year, and expected probability t for getting the profit as parameters, and returns the number of share to buy at the current price (ignore the transaction fees and assume there is no annual dividends).

    (b) Apply this function to the following case, where the annual rate of return has a mean of 0.12 and a standard deviation of 0.06. The current price of the stock is \$35 per share. The client expects to make at least \$1000 profit with a probability of at least 90% in one year. How many shares should the client buy now?

2. [20] Write a Python function testmean(L, t, α), where L is a sample of numbers, t a threshold, and α a significance level, and performs a t-test on whether the mean of L is greater than t. It should print the test result if the α is not null. Otherwise, it should print the p-value with 4 digits precision. Apply this function to answer the following questions.

    (a) Radium-226 is a natural occurring radioactive gas. EPA limit of Radium-226 in soil is 4.0 pCi/L. A sample of elevated levels of radium-226 in a city is as follows (measured in pCi/L)

    ```
    1.46,  0.58,  4.31,  1.02,  0.17,  2.92,  0.91,  0.43,  0.91,  1.30,
    8.24,  3.51,  6.87,  1.43,  1.44,  4.49,  4.21,  1.84,  5.92,  1.86,
    1.41,  1.70,  2.02,  1.65,  1.40,  0.75
    ```

    With a test level $\alpha = 0.1$, can we say that the mean radium-226 level in the sample is greater than the EPA limit?

    (b) Consider another sample from a different city.

    ```
    3.23,  4.16,  4.47,   3.4,  4.02,  4.33,  4.05,  4.14,  4.23,  3.38,
    4.91,  4.02,  4.23,  4.54,  4.14,  4.13,  4.21,  4.66,  4.73,  4.05,
    4.10,  4.11,  4.21,  4.34,  4.22,  4.21,  4.34,  4.35,  4.32,  4.47
    ```

    Find the p-value with 4 digits precision, at which we can say that the sample mean is greater than the EPA limit.

3. [20] The file MAINELAKE.XLS contains mercury levels of 120 lakes in Maine obtained in an environmental study. You should perform the following tasks using this data. In the following, let ELEVATION be the independent variable (X) and MERCURY be the dependent variable (y).

   (a) Plot the histograms for attributes ELEVATION and MERCURY.

   (b) Plot a scatterplot of ELEVATION and MERCURY.

   (c) Use seaborn to plot a linear regression line

   (d) Use statsmodels to obtain a simple linear model (i.e., OLS) $E(y) = \beta_0 + \beta_1 x$, where y is the mercury level and x is the elevation. Print the model summary, and find the values of β_0 and β_1.

   (e) Repeat this using sklearn.